CATISSUE SUITE→ERNE WRAPPER

Data Element Mapping

Presented by Sean Kelly
EDRN Informatics Center
Jet Propulsion Laboratory

ERNE CDES

- 26 in total

- Categories

  - Specimen characteristics

  - Participant information

    - Demographics as well as 1st degree relatives
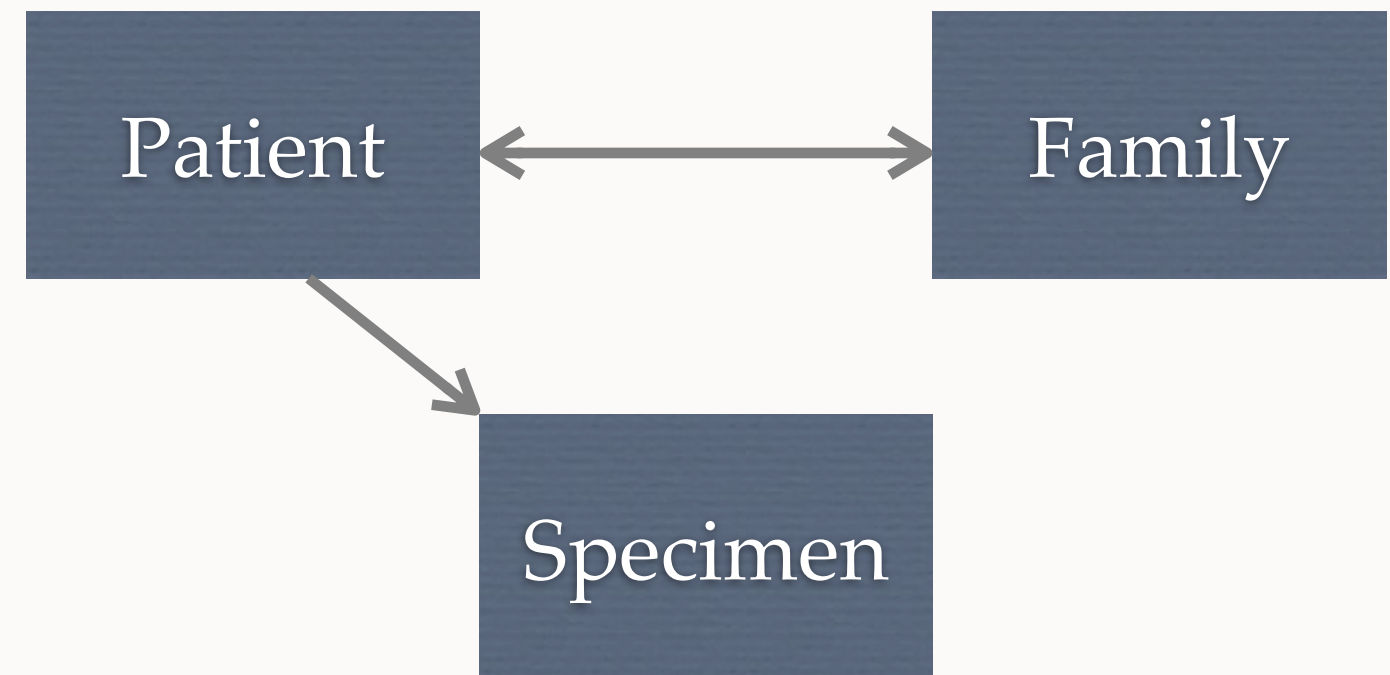
  - Site-specific details

ERNE CDES

- Every CDE may appear in a query constraint

- Every CDE may be a returned value

ERNE DATA

- Relational model

- But ERNE query results hide this
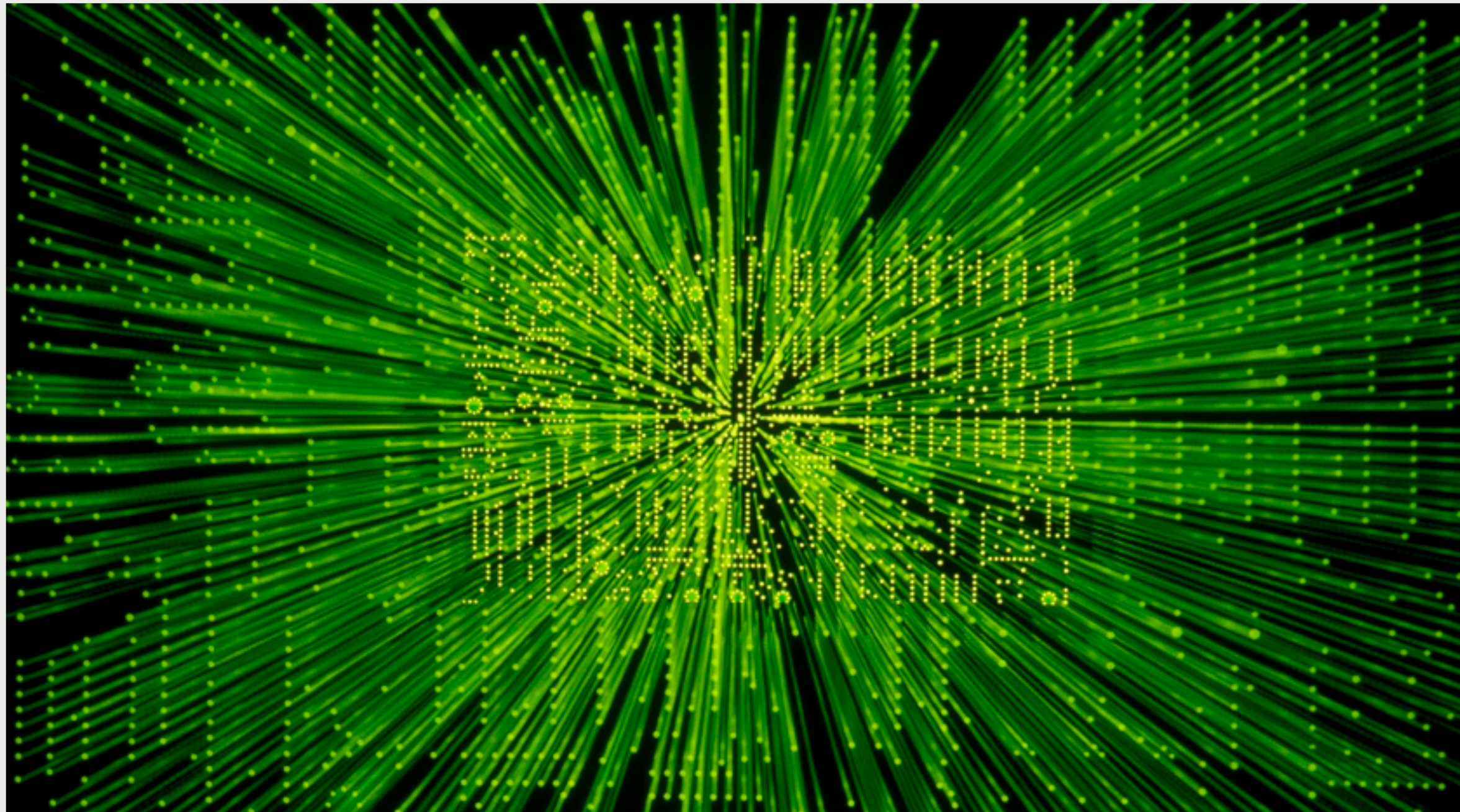
ONE ROW PER MATCHING SPECIMEN

CATISSUE SUITE

- Object oriented

  - Class CATISSUE_SPECIMEN

    - Subclasses CATISSUE_FLUID_SPECIMEN, _CELL_SPECIMEN, etc.

  - Class CATISSUE_PARTICIPANT

  - Etc.

MAPPING PROCESS

- For each ERNE CDE:

  - Identify corresponding field(s) in class(es) in caTissue Suite model

  - Create two functions:

    - To map a query constraint from ERNE to caTissue Suite

    - To map a return value from caTissue Suite to ERNE

THE 26 CDES

In software order

CDE #1

- BASELINE_CANCER-AGE-DIAGNOSIS_VALUE

  - Age of patient at diagnosis, if any

- CATISSUE_ABS_SPECI_COLL_GROUP

  - clinicalDiagnosis

    - A string that may or may not mention patient's age

- Mapping quality: poor

CDE #2

- BASELINE_CANCER-CONFIRMATION_CODE

  - Does patient have cancer? 0 = no, 1 = yes, 9 = unknown/refused

- CATISSUE_ABS_SPECI_COLL_GROUP

  - clinicalDiagnosis

    - A string that names a diagnosis

- Mapping quality: OK

CDE #3

- BASELINE_CANCER-ICD9-CODE

  - Medical industry standard code

- CATISSUE_ABS_SPECI_COLL_GROUP

  - clinicalDiagnosis

    - A string that names a diagnosis; possible mapping to ICD9, possible mention of ICD9

- Mapping quality: poor

CDE #4

- BASELINE_DEMOGRAPHICS-ETHNIC_CODE

  - Hispanic or latino designation: 0=no, 1=yes, 9=unknown/refused

- CATISSUE_PARTICIPANT

  - ethnicity

    - String, but from a controlled vocabulary

- Mapping quality: good

CDE #5

- BASELINE_DEMOGRAPHICS-GENDER_CODE

  - Gender codes, 1=Male, 2=Female, 9=Unknown/refused

- CATISSUE_PARTICIPANT

  - gender

    - Strings from a controlled vocabulary

- Mapping quality: good

CDE #6

- BASELINE_DEMOGRAPHICS_RACE_CODE

  - Race codes, 1=White, 2=Black or African-American, etc.

- CATISSUE_PARTICIPANT

  - race

    - Strings from a controlled vocabulary

- Mapping quality: good

CDE #7

- BASELINE_FAMILY_CANCER-CONFIRMATION_CODE

  - Have any of the participant's living and deceased first degree blood relatives been diagnosed with cancer? 0=no, 1=yes, 9=unknown/refused

- No direct mapping in caTissue Suite 1.1

  - Site-specific forms may capture this

- Mapping quality: null

CDE #8

- BASELINE_FAMILY_CANCER-LOCATION_CODE

  - Cancer type/location for each of the participant's living and deceased first degree blood relatives, 1=bladder, 2=bone, 3=brain, etc.

- No direct mapping in caTissue Suite 1.1

  - Site-specific forms may capture this

- Mapping quality: null

CDE #9

- BASELINE_SMOKE-REGULAR_1YEAR_CODE

  - Did participant smoke cigarettes regularly, at least one a day for a year or more? 0=no, 1=yes, 9=unknown/refused

- No direct mapping in caTissue Suite 1.1

  - Site-specific forms may capture this

- Mapping quality: null

CDE #10

- SPECIMEN_AGE-COLLECTED_VALUE

  - Age of participant when specimen was collected

- CATISSUE_COLL_PROT_EVENT (as well as Participant and CollProtReg)

  - studyCalendarEventPoint

    - DOB of participant/registrationDate/studyCalendarEventPoint

- Mapping quality: good

CDE #11

- SPECIMEN_AMOUNT-STORED_UNIT_CODE

  - Units on the amount of specimen originally stored; 1=mcl, 2=ml, 3=mcg, etc.

- CATISSUE_SPECIMEN subclass

  - Depends on actual subclass; for example, FluidSpecimen will use ml, TissueSpecimen: mcg, etc.

- Mapping quality: good

CDE #12

- SPECIMEN_AMOUNT-STORED_VALUE

  - Amount originally stored, with units depending on unit code

- CATISSUE_SPECIMEN

  - initialQuantity

- Mapping quality: good

CDE #13

- SPECIMEN_AMOUNT_REMAINING_UNIT_CODE

  - Units on the amount of specimen originally stored; 1=mcl, 2=ml, 3=mcg, etc. Should *never* be different from _STORED_UNIT_CODE!

- CATISSUE_SPECIMEN

  - Depends on actual subclass; for example, FluidSpecimen will use ml, TissueSpecimen: mcg, etc.

- Mapping quality: good

CDE #14

- SPECIMEN_AMOUNT_REMAINING_VALUE

  - Amount left, with units depending on unit code

- CATISSUE_SPECIMEN

  - availableQuantity

- Mapping quality: good

CDE #15

- SPECIMEN_AVAILABLE_CODE

  - Is this specimen available? 0=no, 1=yes, 9=unknown/refused

- CATISSUE_SPECIMEN

  - available, boolean

- Mapping quality: good

CDE #16

- SPECIMEN_COLLECTED_CODE

  - What was collected? 1=ascites, 2=biliary washing/brushing, 3=blood, etc.

- CATISSUE_SPECIMEN subclass

  - Actual kind of subclass maps to a code

- Mapping quality: good

CDE #17

- SPECIMEN_CONTACT-EMAIL_TEXT

  - Whom to ask for a shipment of specimens

- CATISSUE_SITE

  - emailAddress, address of responsible person at site

- Mapping quality: good

CDE #18

- SPECIMEN_DAY-CA-DIAGNOSIS_VALUE

  - Number of days participant was diagnosed with cancer from specimen collection day (–, collection before diagnosis; +, collection after diagnosis)

- CATISSUE_ABS_SPECI_COLL_GROUP

  - clinicalDiagnosis

    - A string that may or may not mention diagnosis date

- Mapping quality: poor

CDE #19

- SPECIMEN_FINAL-STORE_CODE

  - Specimen storage temperature; 1=liquid nitrogen, 3=–70ºC/–80ºC, etc.

- CATISSUE_FROZEN_EVENT_PARAM

  - method

    - String describing storage method from controlled vocabulary

- Mapping quality: good

CDE #20

- SPECIMEN_STORED_CODE

  - Type of specimen stored; 1=whole blood, 2=serum, 3=plasma, etc.

- CATISSUE_SPECIMEN_TYPE

  - specimenType

    - Hierarchical vocabulary

- Mapping quality: good

CDE #21

- SPECIMEN_STUDY-DESIGN_CODE

  - What kind of protocol the specimen came from; 1=cohort, 2=case/control, 5=randomized, 15=prospective, etc.

- No direct mapping in caTissue Suite 1.1

  - Perhaps should be configurable on a per-installation basis?

- Mapping quality: null

CDE #22

- SPECIMEN_TISSUE_DEGREE-INVASIVE_CODE

  - Degree to which neoplasm infiltrates surrounding tissue; 1=invasive, 3=pre-invasive, 4=hyperplasia, etc.

- CATISSUE_ABS_SPECI_COLL_GROUP

  - clinicalDiagnosis

    - A string that may or may not include terms like "hyperplasia"

- Mapping quality: poor

CDE #23

- SPECIMEN_TISSUE_ORGAN-SITE_CODE

  - Code identifying where specimen came from; 1=bladder, 2=bone, 3=brain, 4=breast, etc.

- CATISSUE_SPECIMEN_CHAR

  - tissueSite

    - String naming anatomical site from controlled vocabulary

- Mapping quality: good

CDE #24

- STUDY_PARTICIPANT_ID

  - De-identified unique identifier for participant

- CATISSUE_PARTICIPANT

  - id

    - System-assigned integer ID

- Mapping quality: good

CDE #25

- STUDY_PROTOCOL_ID

  - EDRN-assigned identifier

- CATISSUE_SPECIMEN_PROTOCOL

  - id

    - System-assigned integer ID, perhaps mapped to EDRN protocol ID

- Mapping quality: good

CDE #26

- STUDY_SITE_ID

  - Unique EDRN-assigned integer ID for site

- CATISSUE_SITE

  - id

    - System-assigned unique integer ID, perhaps mapped to EDRN site ID

- Mapping quality: good

SITE-SPECIFICS

A new feature of caTissue Suite 1.1

CATISSUE SUITE 1.1

- New feature: Local Extensions

  - Build-your-own forms on the fly

    - Through-the-web form editor

  - Enables capture of custom data elements

    - Perhaps smoking history, first-degree relatives cancer, or future ERNE CDEs

WRAPPER RAMIFICATIONS

- Wrapper should be configurable

  - Query locally-added form data

  - Return annotations converted to ERNE format

CONVERTING LOCAL DATA

- Mapping functions

  - From ERNE query constraints and to ERNE CDE results

- How to deploy at each caTissue Suite site?

  - Configuration file?

  - Plug-in code modules?

REFERENCES

Links to useful materials

- caTissue Suite home page: https://cabig.nci.nih.gov/tools/catissuesuite

- caTissue Suite v 1.1.2 tool page: http://bit.ly/9HzPXJ

- caTissue G-Forge: http://gforge.nci.nih.gov/projects/cacorecvs/

- Initial ERNE/caTissue Core mapping: http://bit.ly/cmmPPQ